

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 10 - 307599

(43) 公開日 平成 10 年 (1998) 11 月 17 日

(51) Int. Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G10L 9/14			G10L 9/14	M
				J
9/00			9/00	M
H03M 7/30			H03M 7/30	B

審査請求 未請求 請求項の数 24 O L (全 16 頁)

(21) 出願番号 特願平 10 - 57604
(22) 出願日 平成 10 年 (1998) 3 月 10 日
(31) 優先権主張番号 08 / 814075
(32) 優先日 1997 年 3 月 10 日
(33) 優先権主張国 米国 (U S)

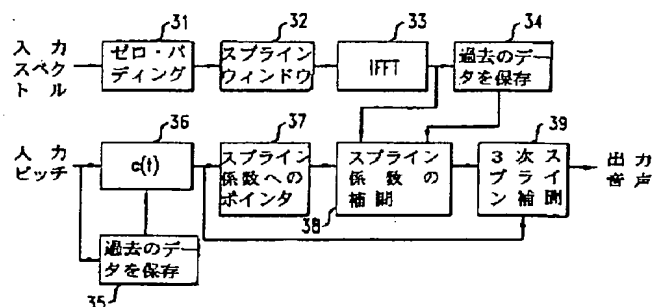
(71) 出願人 596092698
ルーセント テクノロジーズ インコーポ
レーテッド
アメリカ合衆国 07974-0636
ニュージャージー、マレイ ヒル、マウン
テン アヴェニュー 600
(72) 発明者 ヤイア ショーハム
アメリカ合衆国 07060 ニュージャ
ーシー、ウォッチング、ジョンストン ド
ライヴ 645
(74) 代理人 弁理士 岡部 正夫 (外 11 名)

(54) 【発明の名称】 スプラインを使用する波形補間音声コーディング

(57) 【要約】

【課題】 本発明は、受信波形間の補間がそれに基づいて生成されるスプライン係数を使用して行なわれるような低ビットレート波形補間音声デコーダにおける波形補間を行うための複雑さの低い方法と装置に関する。

【解決手段】 各々が対応するピッチ周期の音声信号セグメントを表す周波数領域パラメータの集合を含む 2 つの信号が波形補間エンコーダから受信される。その後、受信された各信号からスプライン係数が生成されるが、その際スプライン係数の各集合は、周波数領域パラメータの対応する集合の時間領域変換のスプライン表示を含む。最後に、デコーダがスプライン表示間を補間し、再生音声信号を合成するために使用される補間された時間領域データを生成する。本発明のある実施形態では、少なくとも 1 つのスプライン表示の時間スケールが修正され、相互の補間を可能にする。また、本発明の 1 つの例示としての実施形態の 1 つによれば、3 次スプライン表示が使用されるが、他の例示としての実施形態によれば、計数論的スプライン表示の新しい変形が有利にも利用される。



【特許請求の範囲】

【請求項 1】 通信チャネルを経由して通信されたエンコーディングされた信号に基づいて再生音声信号を合成する方法であって、該方法が、

第 1 のピッチ周期と等しい長さの第 1 の音声信号セグメントを表す周波数領域パラメータの第 1 の集合を含む第 1 の通信信号と、第 2 のピッチ周期と等しい長さの第 2 の音声信号セグメントを表す周波数領域パラメータの第 2 の集合を含む第 2 の通信信号とを含む少なくとも 2 つの通信信号を受信するステップと、

周波数領域パラメータの前記第 1 の集合の時間領域変換のスプライン表示を含むスプライン係数の第 1 の集合と、周波数領域パラメータの前記第 2 の集合の時間領域変換のスプライン表示を含むスプライン係数の第 2 の集合とを含むスプライン係数の少なくとも 2 つの集合を生成するステップと、

周波数領域パラメータの前記第 1 の集合の前記時間領域変換の前記スプライン表示と、周波数領域パラメータの前記第 2 の集合の前記時間領域変換の前記スプライン表示との間を補間することによって前記再生信号を合成するステップとを含む方法。

【請求項 2】 請求項 1 に記載の方法において、前記スプライン表示が 3 次スプライン表示を含む方法。

【請求項 3】 請求項 1 に記載の方法において、前記スプライン表示が計数論的スプライン表示に基づいている方法。

【請求項 4】 請求項 3 に記載の方法において、前記スプライン表示が有限サポート基底関数を有する方法。

【請求項 5】 請求項 4 に記載の方法において、前記スプライン表示がそれに対応する前記時間領域変換のサンプルを含む方法。

【請求項 6】 請求項 1 に記載の方法において、前記第 1 のピッチ周期と前記第 2 のピッチ周期が等しくなく、前記再生信号を合成するステップが周波数領域パラメータの前記第 2 の集合の前記時間領域変換の少なくとも前記スプライン表示の時間スケールを修正するステップを含む方法。

【請求項 7】 請求項 1 に記載の方法において、さらに、時間領域パラメータの対応する第 1 および第 2 の集合を生じるために、周波数領域パラメータの前記第 1 および第 2 の集合に対して逆変換を行うステップを含み、その際前記生成ステップが時間領域パラメータの前記第 1 および第 2 の集合に基づいている方法。

【請求項 8】 請求項 7 に記載の方法において、さらに、前期逆変換を行うステップの前に、周波数領域パラメータの前記第 1 および第 2 の集合にゼロ・パディングを行い固定した基数 2 のサイズにするステップを含む方法。

【請求項 9】 請求項 8 に記載の方法において、前記逆変換が I F F T から成る方法。

【請求項 1 0】 請求項 1 に記載の方法において、前記再生信号を合成する前記ステップが、

連続時間領域信号のスプライン表示を含む補間されたスプライン係数の集合を生成するステップと、補間されたスプライン係数の前記集合に基づいて前記再生信号を生成するステップとを含む方法。

【請求項 1 1】 請求項 1 0 に記載の方法において、前記再生信号は、不均一なレートで前記連続時間領域信号をサンプリングすることによって生成される方法。

10 【請求項 1 2】 請求項 1 1 に記載の方法において、前記不均一なレートが前記第 1 および第 2 のピッチ周期に基づいて決定される方法。

【請求項 1 3】 通信チャネルを経由して通信されるエンコーディングされた信号に基づいて再生音声信号を合成する音声デコーダであって、該デコーダが、

第 1 のピッチ周期に等しい長さの第 1 の音声信号セグメントを表す周波数領域パラメータの第 1 の集合を含む第 1 の通信信号と、第 2 のピッチ周期に等しい長さの第 2 の音声信号セグメントを表す周波数領域パラメータの第 2 の集合を含む第 2 の通信信号とを含む少なくとも 2 つの通信信号を受信する信号受信機と、

周波数領域パラメータの前記第 1 の集合の時間領域変換のスプライン表示を含むスプライン係数の第 1 の集合と、周波数領域パラメータの前記第 2 の集合の時間領域変換のスプライン表示を含むスプライン係数の第 2 の集合とを含む、スプライン係数の少なくとも 2 つの集合を生成するスプライン係数発生器と、

周波数領域パラメータの前記第 1 の集合の前記時間領域変換の前記スプライン表示と、周波数領域パラメータの前記第 2 の集合の前記時間領域変換の前記スプライン表示との間を補間することによって前記再生信号を合成する信号シンセサイザとを含むデコーダ。

【請求項 1 4】 請求項 1 3 に記載のデコーダにおいて、前記スプライン表示が 3 次スプライン表示を含むデコーダ。

【請求項 1 5】 請求項 1 3 に記載のデコーダにおいて、前記スプライン表示が計数論的スプライン表示に基づいているデコーダ。

【請求項 1 6】 請求項 1 5 に記載のデコーダにおいて、前記スプライン表示が有限サポート基底関数を有するデコーダ。

【請求項 1 7】 請求項 1 6 に記載のデコーダにおいて、前記スプライン表示がそれに対応する前記時間領域変換のサンプルを含むデコーダ。

【請求項 1 8】 請求項 1 3 に記載のデコーダにおいて、前記第 1 のピッチ周期と前記第 2 のピッチ周期が等しくなく、前記信号シンセサイザが周波数領域パラメータの前記第 2 の集合の前記時間領域変換の少なくとも前記スプライン表示の前記時間スケールを修正するための手段を含むデコーダ。

【請求項 1 9】 請求項 1 3 に記載のデコーダにおいて、さらに、時間領域パラメータの対応する第 1 および第 2 の集合を生じるために、周波数領域パラメータの前記第 1 および第 2 の集合に対して行われる逆変換を含み、その際前記スプライン係数発生器が時間領域パラメータの前記第 1 および第 2 の集合に基づいているデコーダ。

【請求項 2 0】 請求項 1 9 に記載のデコーダにおいて、さらに、前期逆変換で使用するために、周波数領域パラメータの前記第 1 および第 2 の集合をゼロ・パディングして固定した基数 2 のサイズにするための手段を含むデコーダ。

【請求項 2 1】 請求項 2 0 に記載のデコーダにおいて、前期逆変換が I F F T から成るデコーダ。

【請求項 2 2】 請求項 1 3 に記載のデコーダにおいて、前記信号シンセサイザが、連続時間領域信号のスプライン表示を含む補間スプライン係数の集合を生成するための手段と、補間スプライン係数の前記集合に基づいて前記再生信号を生成するための手段を含むデコーダ。

【請求項 2 3】 請求項 2 2 に記載のデコーダにおいて、前記再生信号が不均一なレートで前記連続時間領域信号をサンプリングすることによって生成されるデコーダ。

【請求項 2 4】 請求項 2 3 に記載のデコーダにおいて、前記不均一なレートが前記第 1 および第 2 のピッチ周期に基づいて決定されるデコーダ。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】 本発明は、概して、低ビットレート音声コーディングの分野に関し、より詳細には、複雑さを低減して低ビットレート音声コーディングを行うための方法と装置に関する。

【 0 0 0 3 】

【従来の技術、及び、発明が解決しようとする課題】 音声情報の通信は、音声を表す電気信号をあるチャネルまたはネットワーク（「チャネル」）上で伝送することを含むことが多い。音声通信において一般に遭遇する問題は、制限された容量または帯域幅のチャネルを通じて音声を伝送する方法である（現代のデジタル通信システムでは、帯域幅はビットレートによって表されることが多い）。制限されたチャネル帯域幅の問題は普通、音声信号を圧縮してチャネル帯域幅の要求に合致させる音声コーディング・システムの適用によって扱われる。音声コーディング・システムには、音声信号をチャネル上で伝送するためのコード語に変換するエンコーダと、受信されたコード語から音声を再生するデコーダとが含まれる。

【 0 0 0 4 】 一般的な問題として、信号圧縮を伴う大部分の音声コーディング・システムの目的は、有声音声の

ような原音声の忠実な再生である。有声音声は、話者の声帯が緊張し、準周期的に振動するとき発生する。時間領域では、有声音声信号は、ピッチサイクルと呼ばれる一連の同様な波形がゆっくりと変化する波形として現れる。各ピッチサイクルは、ピッチ周期と呼ばれる周期を有する。ピッチサイクル波形と同様、ピッチ周期は、あるピッチサイクルから次のピッチサイクルにゆっくりと変化する。

【 0 0 0 5 】 約 8 キロビット／秒（k b p s）で動作する音声コーディング・システムの多くは、音声生成プロセスの知識を活用することによって原音声波形をコーディングする。こうしたいわゆる波形コードの例としてコード励起線形予測（C E L P）音声コーディング・システムがあるが、これは音声波形を時間と共に変化する線形予測（L P）フィルタによってフィルタリングして残差音声信号を生じることによってコーディングするものである。有声音声中、残差信号は一連のピッチサイクルを含み、その各々にはピッチパルスと呼ばれる主要な経過音とそれを取り巻く一連の低振幅振動とが含まれる。残差信号は C E L P システムによってコードブックからの基準化された固定長ベクトルの連結として表される。有声音声の高いコーディング効率を達成するために、C E L P の実現例の大部分にはまた、適当な周期で通信された信号の再生を促進する長期予測子（または適応コードブック）も含まれる。しかし、時とともに改善されているにもかかわらず、波形コーディング・システムには、6 k b / s 以下のレートで動作するとき、知覚的に有意な歪みを有するものが多い。この歪みは通常雑音として特徴づけられる。

【 0 0 0 6 】 例えば 2 . 4 k b / s のレートで動作する低ビットレートコーディング・システムは一般にその性質上パラメトリックである。すなわち、それらは定期的な間隔で、音声信号のピッチ周期と空間エンベロップ（またはホルマント）を記述するパラメータを伝送することによって動作する。こうしたいわゆるパラメトリック・コードの例が L P ボコーダ・システムである。L P ボコーダはピッチ周期毎に 1 つのパルスによって有声音声信号をモデル化する。この基本技術は、とりわけ空間エンベロップに関する伝送情報を含むよう増補される。L P ボコーダは一般に妥当な性能を提供するが、やはり通常ブザー音として特徴づけられる知覚的に有意な歪みを導入する。

【 0 0 0 7 】 上記で論じた種類および他の正弦コーディング・システムで一般的な「反響」は、一般に、再生された音声信号に原有声音声にあるピッチサイクルの強弱が（全体として、またはかなりの部分で）欠けていることの結果である。当然、こうした種類の歪みは、低いビットレートでは音声の強弱に関する情報をコーディングする音声コーディング・システムの能力が低下するためより多く出現する。こうした問題は、波形補間および関

連する信号モデル化技術に基づくアルゴリズムの導入に伴って扱われ、最近低レート音声コーディングにおける大きな進歩を達成した。こうした技術の背後にある一般的な概念は、原音声に関して送信される情報をできる限り少なくする一方で、原音声の自然の変化を模倣するコーディングされた信号を合成することである。この概念は、音声は普通低レートでサンプリングおよび補間可能なゆっくり変化する属性を伝えるという観測に基づいている。信号中のかなりの量の情報は、ある重要な特徴が忠実に再生される限り捨てることができる。

【0008】このタスクを達成する際使用される主要な技術は波形補間(WI)と信号分解(SD)である。WIは合成プロセスで(すなわちデコーダで)使用され、普通音声信号、特に有声領域で観測される平滑さの度合いを維持する。平滑さを維持することによってコーディング歪みに対する頑強性が向上する。一例として、ピッチが不意に(不自然に)でなく平滑に変化する場合、より大きなピッチのエラーが知覚的に許容できる。同じことが他の種類の歪みについても言える。SDによってコーディング・システムがより重要な信号領域に焦点を当て、あまり重要でない領域で伝えられる情報を捨てることが可能になる。WIコードは、例えば、Y. Shoham「時間一周波数補間に基づく2.4~4.0 kbpsでの高品質音声コーディング」、ICASSP'93会報、II167~170ページ、Y. Shoham「時間一周波数補間に基づく2.4 kbpsでの高品質音声コーディング」、Eurospeech'93会報、741~744ページ、W. B. Kleijn他「特性波形の分解に基づく音声コード」、ICASSP'95会報、508~511ページおよびW. B. Kleijn他「複雑さの低い補間コード」、ICASSP'96会報、212~215ページに説明されている。WIコードはまた、上記で援用した共通に譲渡された米国特許出願「プロトタイプ波形音声コーディングのための方法と装置」第08/667,295号および1996年5月14日W. B. Kleijnに対して発行された共通所有の米国特許第5,517,595号でも説明されるが、この特許は、ここに記載されるのと同様に、引用によって本明細書の記載に援用する。

【0009】WIコードは一般に低ビットレートで適度に良好な品質の再生された音声を生じるが、こうした従来技術のコードは、例えば、低コスト端末での使用を商業的に実現可能にするには高価すぎるが多い。従って、十分な性能レベルを維持しつつ(すなわち、再生された音声の質に関して)従来技術のWIコードより実質上低い複雑さを有するWIコードが入手可能になることが望ましい。

【0010】

【課題を解決するための手段】本発明によれば、低ビットレートWI音声デコーダで波形補間を行うための改善

された方法と装置が提供されるが、その際受信波形間の補間はそれに基づいて生成されたスプライン係数を使用して行われる。詳細には、各々が対応するピッチ周期の音声信号セグメントを表す周波数領域パラメータの集合を含む2つの信号がWIエンコーダから受信される。その後、受信された各信号からスプライン係数が生成されるが、その際スプライン係数の各集合は周波数領域パラメータの対応する集合の時間領域変換のスプライン表示を含む。そして、最後に、デコーダはスプライン表示の間を補間し、再生音声信号を合成するために使用される補間時間領域データを生成する。本発明のある実施形態では、少なくとも1つのスプライン表示の時間スケールが修正され、相互の補間を可能にする。また、本発明の1つの例示としての実施形態によれば、3次スプライン表示が使用されるが、他の例示としての実施形態によれば、新しい種類の係数論のスプライン表示が有利にも利用される。

【0011】

【発明の実施の形態】

< A. 波形補間の概観 > WI法はスペクトルの時間シーケンスの処理に基づいている。このシーケンスにおけるスペクトルは、例えば、音声信号のピッチ長スナップショットの位相緩和離散フーリエ変換(DFT)である。さらに、スペクトルの位相は循環シフトの対象となる。スナップショットは、原則として1つのサンプルと同じ短いアップデート間隔に得られる。こうしたアップデート間隔は全体としてピッチに依存するが、有効な処理のために、好適にはピッチ周期に動的に適合される。

【0012】WIプロセスが以下例示として説明される。S(t, K)は時間tにおけるスナップショットのDFTであり、ピッチ周期P(t)は時間とともに変化する。U(t, c)によって示されるS(t, K)の逆DFT(IDFT)が、T秒の大きさの一定のDFT基底関数サポートに関連して得られる。これは時間スケール正規化として知られ、当業技術分野に熟練した者には周知である。この正規化によって、U(t, c)は周期Tを有する軸cに沿った周期関数として見られる。2つの連続するスナップショットをt₁とt₂で取ると、S(t₁, K)は、最大相関に関する循環シフトによってS(t₂, K)と整合するので有利である。従って、ピッチ信号がゆっくり変化する場合、2次元曲面U(t, c)はt軸に沿って平滑である。この状態は図1に例示として示されているが、ここではすべての波形はc軸に沿って同じ周期Tを有し、t軸に沿ってゆっくり変化する。実際には、曲面U(t, c)は何らかの特定の点ではなく、スペクトルS(t₁, K)およびS(t₂, K)に対応する境界波形U(t₁, c)およびU(t₂, c)で与えられる。中間の数値は、以下説明されるように、有利にもこれらのスペクトルから補間される。U(t, c)における変数「c」は正規化ピッチサ

イクルの数を示す。音声信号の場合、これは、 $c(t)$ によって示される時間の関数であり、次の式によって与

$$c(t) = T \int^t \frac{1}{P(v)} dv$$

時間 t におけるサイクルの数値を与えると、点 $(t, c(t))$ で曲面をサンプリングすることによって1次元信号 $s(t)$ が生成される。すなわち、

$$s(t) = U(t, c(t)) \quad (2)$$

図1に例示されるように、 $s(t)$ は、 $c(t)$ によって定義される経路に沿った、すなわち位置 $(t, c(t))$ にある $U(t, c)$ をサンプリングすることに

$$s(t) = U(t, c(t)) = \sum_K S(t, K) e^{j \frac{2\pi K}{T} c(t)} \quad (3)$$

ここでスペクトル $S(t, K)$ は次の2つの境界スペク

$$S(t, K) = \alpha(t) S(t_0, K) + \beta(t) S(t_1, K) \quad t_0 < t < t_1 \quad (4)$$

関数 $\alpha(t)$ および $\beta(t)$ は、例えば、線形補間を表すが、他の補間規則、詳細にはスペクトルの振幅と位相を別々に補間するものなどが代替的に利用されることもある。サイクル関数 $c(t)$ も補間によって得られるので有利である。第1に、ピッチ関数 $P(t)$ がその境界数値 $P(t_0)$ および $P(t_1)$ から補間され、その後上記の式(1)が $t_0 < t < t_1$ について計算される。

【0013】アップデート・スペクトルの忠実な伝送を仮定すると、信号 $s(t)$ は原音声の重要な特性の大部分を有する。詳細には、そのピッチ・トラックは、たとえピッチの同期が使用されず、アップデート時間がピッチと無関係であっても原音声のそれに従っている。これは低レート・コーディングのために有利な大量の情報の低減を意味する。

【0014】非周期的(無声)音声セグメントでは、ピッチは本質的に任意の数値がエンコードのピッチ検出器によって計算されるものすべてに設定されるので、真のピッチ・サイクルを表さない。さらに、結果として生じるピッチの数値はピッチ・トラックを平滑にするために修正されるので有利である。こうしたピッチは、その真の性質に関わらず、同じ方法でシステムによって使用される。このアプローチは有利にも音声化の分類を除去し、頑強な処理を提供する。この場合でも、(実際はすべての信号について)上記で説明した補間の枠組みは、アップデート間隔がピッチ周期の半分未満の時はいつても良好に機能することに留意されたい。

【0015】 W I コードにおける信号分解の概観 > W I エンコードは通常有効な圧縮のために音声信号を分析し分解する。詳細には、信号分解は2つのレベルで行われるので有利である。第1のレベルでは、標準10次 L P C 分析が、例えば、25 msec のフレームにわたるフレーム毎に一度行われ、スペクトル・エンベロー

$$S(t, K) = A(t, K) R(t, K) \quad (5)$$

ここで $A(t, K)$ は時間 t の L P C スペクトルで 50 ある。

えられる。

【数1】

(1)

よって生成される。完全な曲面 $U(t, c)$ は、図1に例示としての目的のためにだけ示される。実際には、普通サンプリングの前に曲面全体を生成(すなわち補間)する必要はない。サンプリング経路 $(t, c(t))$ のこれらの数値だけが有利にも次の計算によって決定される。

【数2】

トルから補間される。

ブ(L P C) パラメータおよび L P 残差信号を得る。この方法による信号の分割はスペクトルの知覚的に有効な量子化を考慮している。高品質の再生音声を生じるためにはスペクトル・エンベロープの適度に正確なコーディングが好適であるが、特に高い周波数では、微細構造の L P 残差スペクトルのかなりの歪みが許容されることが多い。このことを考慮して、残差信号は有利にも第2レベルの分解を経験するが、その目的は信号を構造化成分と非構造化成分とに分解することである。構造化信号は本質的に周期的であるが、非構造化成分は非周期的で本質的にランダム(すなわち雑音状)である。

【0016】高度低レート音声コードでは、様々な方法と手順のこの種の基本的分解を使用するものが多いが、大部分の W I コードでは、2次レベル分解は、ゆっくり変化する波形(S E W)と急速に変化する波形(R E W)の概念を使用して行われる(例えば、各々上記で援用された W. B. K l e i j n 他「特性波形の分解に基づく音声コード」および米国特許第5, 517, 595号を参照されたい)。このアプローチは、有声の(すなわち大部分周期的な)音声セグメントでは、ピッチやスペクトル・パラメータのような音響的特徴は比較的ゆっくり変化するが、無声セグメントではこれらの特徴はより早く変化するという観測に基づいている。従って、信号が S E W 成分と R E W 成分に分割される場合、S E W は大部分周期的な成分を表し、R E W 成分は大部分非周期的な雑音状信号を表す。この分解は L P 残差領域で行われるので有利である。この目的のために、残差のアップデート・スナップショットが、時間 t のピッチサイズ D F T を取り、それによってスペクトル $R(t, K)$ を生じることによって得られる。従って、この音声スペクトルは、以下の式によって与えられる。

【0017】SEWシーケンスは、例えば、20Hz、20タップ・ローパスフィルタを使用して、時間軸に沿って $R(t, K)$ の各スペクトル成分を（すなわち K の各数値について）フィルタリングすることによって得られる。これからSEWスペクトルのシーケンス、 $SEW(t, K)$ が生じるが、これは、例えば、1フレーム当たり1SEWスペクトルにダウンサンプリングされるので有利である。相補ハイパスフィルタを使用することによって、REWスペクトルのシーケンス、 $REW(t, K)$ が同様に得られる。スペクトル・スナップショットは普通正確なピッチサイクル間隔では得られないので、スペクトル $S(t, K)$ はフィルタリングの前に整合されるので有利である。この整合は、例えば、時間領域循環シフトと同等の高解像度位相整合を含み、現在と以前とのスペクトルの相関を最大にするので有利である。これは位相の不整合による人工的なスペクトルの変化を除去する。

【0018】1つの興味深い観測は、多くの他の分解方法と異なって、この分解は（少なくとも原理的には）損失がなく可逆的である、すなわちもとの（整列した）シーケンス $R(t, K)$ が回復できるということである。従って、この方法ではコーディング性能に上限がない。SEWとREWが十分高いビットレートでコーディングされれば、（完全な残差信号が正確に再生されるので）従来のWIデコーダによって非常に高品質の音声が生じ得る。

【0019】スペクトル $R(t, K)$ は有利にも正規化され、 K 軸について単位平均2乗平均平方根（RMS）値を有する。これはレベルの変動を除去し、 SEW/REW 分析を向上させ、REWとSEWの量子化を容易にする。RMSレベル（すなわち利得）は個別に量子化されることもある。これはまたシステムが他のパラメータと無関係に、信号レベルの知覚的に重要な変化（例えば、有声音の開始）に特別な注意を払うようにする。

【0020】<C. 従来の波形補間コード>図2は、エンコーダ21およびデコーダ22を含む従来のWIコードのブロック図を示す。エンコーダでは、LP分析（ブロック212）が入力音声に適用され、LPフィルタが使用されてLP残差が得られる（ブロック211）。ピッチ推定器214が残差に適用され現在のピッチ周期が得られる。ピッチサイズ・スナップショット（ブロック213）が残差について取られ、DFTによって変換され正規化される（ブロック215）。結果として生じるスペクトルのシーケンスはまず整合され（ブロック217）、時間軸に沿ってフィルタリングされてSEW（ブロック218）およびREW（ブロック219）信号を形成する。これらは量子化され、（ブロック212で生成する）ピッチLP係数および（ブロック216で生成する）スペクトル利得とともに伝送される。

【0021】デコーダでは、コーディングされたREWおよびSEW信号がデコードされ結合（ブロック223）されて量子化励起スペクトル $R(t, K)$ を形成する。スペクトルはその後LPCスペクトルによって再形成され、利得によって適切なRMSレベルに再基準化され（ブロック222）、それによって量子化音声スペクトル $S(t, K)$ を生じる。ここでこのスペクトルは上記で説明したように補間され（ブロック224）、最終的な再生信号を形成する。

【0022】図2のWIコードは、すべてのデータ、特にREWおよびSEW信号について十分なビット・リソースが利用可能である限り、高品質音声を提供することができる。2つのフルサイズのスペクトルが表示されているので、 REW/SEW 表示は、原則として、オーバーサンプリングされていることに留意されたい。これは量子化器に余分の負担をかける。低ビットレートでは、以下さらに説明するように、ビットが不足しているので、 REW/SEW 表示は意味のある量子化を考慮して、通常大きく妥協される。例えば、2.4kbpsで動作する通常の従来のWIコードは、25msecのフレームサイズを使用するので、通常LPCデータに30ビット、ピッチ情報に7ビット、SEWデータに7ビット、REWデータに6ビットおよび利得情報に10ビットからなるビット割当てを利用するよう制限されている。同様に、1.2kbpsのレートで動作する通常の従来のWIコードは37.5msecのフレームサイズを使用するので、通常LPCデータに25ビット、ピッチ情報に7ビット、SEWデータにビットなし、REWデータに5ビットおよび利得情報に7ビットからなるビット割当てを利用するよう制限されている（1.2kbpsの場合、全体に平滑なLPスペクトルが仮定され、SEW信号は、コーディングされたREW信号部分と相補的なその一部分であると仮定されることに留意されたい）。

【0023】上記で説明した補間コーディングは計算上複雑である。初期のWIコードの中には実際に実時間より動作の遅いものがあった。改良形の複雑さの低いWIコードが上記の「複雑さの低い補間コード」でW. B. Kleijn他によって提案されたが、広範な適用業務範囲で商業的に実行可能な代替案を提供するにはより複雑さの低いコードが必要である。詳細には、例えば、ネットワークのような、他のタスクが中断されずに行えるように、プロセッサの計算能力のわずかな部分だけが使用されるのが望ましい。

【0024】通常のWIコードでは、計算負荷の主な原因は信号分解と補間処理であることに留意されたい。他の重大な原因はピッチ・トラッキング、スペクトル整合およびLPC量子化処理である。安価な実現を達成しようとする場合、メモリの使用も重要な要素である。通常の従来の技術WIコードは時間フィルタリング等の演算に

関するREWとSEWのシーケンスを保持するために多量のRAMを必要とする。全体として、通常の従来のW Iコードでは約6KワードのRAMが必要である。さらに、LPC量子化のために多量のROM(通常約11Kワード)が必要である。

【0025】<D>3次スプラインを使用する複雑さの低い波形補間>従来のW Iコードで行われる、上記で説明したような波形補間処理は、部分的には、時間インスタンス毎に全スペクトル・ベクトルを補間する必要があり、DFT形演算(例えば、上記の(3)の式の計算) 10 を実行する必要があるために非常に複雑である。式

(3)によって実行される三角関数の非正則サンプリングでは、この関数を実行するために有益な簡単な帰納的方法がないためそれは一層複雑になっている。この問題を解決するために、波形補間処理は、本発明の例示としての実施形態に従って、以下のようなより簡単な方法に近づけることができるので有利である。スペクトルS

(1, ..., K)はまずゼロ・パディングによって固定した基数2のサイズに増加される。逆高速フーリエ変換(IFFT)がアップデート毎に一度行われ、固定サイズT 20 の時間信号が得られる。これらの信号は3次スプライン係数ベクトルに変換される(以下より完全に説明される3次スプライン係数は信号処理技術に熟練した者には周知である)。このスプライン係数を使用すると、任意の

$$s(t) = \sum_{n=-\infty}^{\infty} q_n B_k(t-n) \quad (6)$$

ここで q_n はスプライン係数であり、 $B_k(t)$ は、区分的 k 次多項式からなるスプライン連続時間基底関数である。スプライン表示を使用する利点の1つは、基底関数が小さな有限のサポートを有するという事実に見られる。すなわち、サイズ $k+1$ のサポートの時だけ非ゼロである。これは、式(6)の合計は実際には $k+1$ 係数についてだけ行えばよいということであり、従来の帯域制限されたフィルタリングと比較して大きな計算負荷(およびメモリ)の節約を意味する。基底サポートは、 $n=-k+1, \dots, k-1$ の時、 $t=n$ の時点で、ノードと呼ばれる $k+1$ のセクションに分割される。基底は対称的であり、 $B_k(0)=1$ で $B_k(t \geq k-1)=0$ である。従って、 $B_k(t)$ は $(k-1)$ 次多 40 項式を正の $k-1$ セクションに割り当てることによって

$$s(t-n) = [(t-n)^3, (t-n)^2, t-n, 1]$$

ここで $n \leq t \leq n+1$ である。 $s(n)$ がサイズNの離散時間サンプリング・シーケンスであるとし、その基礎となる連続信号 $s(t)$ を推定することが望ましいとす 50

望ましい点の信号の連続時間推定のサンプルが生成されるが、これは有利にも上記の式(1)の関数 $c(t)$ によって決定される動的時間基準化を考慮している。

【0026】信号のスプライン表示の使用は、信号を離散的な時間表示から連続時間表示に変換するための周知の技術である(例えば、M. Unser他「B-スプライン信号処理：第I部-理論」、米国電気電子学会信号処理会報第41巻、第2号、1993年2月、821~833ページ、M. Unser他「B-スプライン信号処理：第II部-有効な設計」、米国電気電子学会信号処理会報第41巻、第2号、1993年2月、834~848ページおよびH. Hou他「画像補間とデジタル・フィルタリングのための3次スプライン」、米国電気電子学会音響信号処理会報第ASSP-26巻、第6号、1978年12月、508~517ページを参照されたい)。帯域制限された信号の場合、それはナイキスト・サンプリングされた数値から連続信号を完全に再生するはるかに高価な無限サポート「サイン(x)/x」フィルタリング演算の代わりに使用できる。

【0027】信号処理技術に熟練した者に周知のように、信号の k 次スプライン表示 $s(t)$ は、次の式のように定義される。

【数3】

完全に定義される。 $(k-1)(k+1)$ 多項式パラメータはノードに連続条件を課すことによって解かれる。すなわち、 $B_k(t)$ の0次から $(k-1)$ 次の導関数はノードで連続的であるので有利である。

【0028】3次スプラインがごく低い計算負荷で大部分の信号の商品品質補間を行うために十分であることは当業技術分野に熟練した者には周知である。従って、本発明の1つの例示としての実施形態によれば、3次スプラインが複雑さの低いW Iコードで波形補間を行う際使用される。上記の定義を $B_k(t)$ (すなわち3次スプライン基底)に適用すると、当業技術分野に熟練した者に明らかなように、式(6)は以下のような行列の形態になる。 40

【数4】

$$\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \\ q_{n+1} \\ q_{n+2} \end{bmatrix} \quad (7)$$

る。上記の式(7)から、 $t=n$ とすると、次の式が導かれる。

$$s(n) = q_{n-1} + 4q_n + q_{n+1}$$

これは、当業技術分野に普通に熟練した者には周知の、IIR（無限インパルス応答）フィルタリング演算の形態の、信号からスプライン係数への変換を定義する。このフィルタは非因果的なので、安定した形態で実現するには留意が必要である。また、2つの初期条件の適切な組み合わせが選択される必要がある。当業技術分野における通常の技術を有する者には周知のように、安定した

$$f_n = p f_{n-1} + s(n); n = 0 \sim N-1$$

$$q_n = p(f_n - q_{n-1}); n = N-1 \sim 0$$

この変換を完全に定義するために、初期値 f_0 と q_0 を知る必要がある。それ自体、本発明の例示としての実施形態の1つによれば、 $f_0 = q_0 = 0$ である。本発明によれば、これらの初期値を割り当てるために本質的に任意の方法が使用できるが、異なった方法は、特に境界近くでは $s(t)$ について異なった数値を生じることに留意されたい。それにも関わらず、結果として生じるさまざまな $s(t)$ は $t = n$ でサンプリングされるとき、同じシーケンス s_n を生じるので有利である。

【0029】本発明の他の例示としての実施形態によれば、初期条件を設定するための別の方法が利用される。この方法は $s(n)$ が周期 N を伴う周期的なものであるという仮定に基づいている。明らかに、これは q_n も周期的であることを意味する。この場合、 $s(n)$ と q_n の関係がDFT演算によって周波数領域で表される場合、初期条件は暗黙のうちに決定され、これに関してそれ以上注意を払う必要はない。また、安定性はこの場合重要でない。

【0030】式(8)に関連するDFT領域フィルタ $H(K)$ は次のシーケンスのDFTを計算することによって得られる。

【数5】

$$h_n = \begin{cases} 4, & n = 0 \\ 1, & n = 1 \\ 1, & n = N-1 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

すなわち、 $H(K) = \text{DFT}\{h_n\}$ である。同様に、 $S(K) = \text{DFT}\{s(n)\}$ であり $Q(K) = \text{DFT}\{q_n\}$ である。従って、式(8)のDFTバージョンは単純に $S(K) = H(K)Q(K)$ である。スプライン・ウィンドウを $W(K) = 1/H(K)$ と定義すると、次のスプライン変換が得られる。

$$Q(K) = W(K)S(K) \quad (11)$$

複合ウィンドウ $W(K)$ は有利にも一度オフラインで計算され、ROMに保存されることに留意されたい。また、変換の複雑さは単に入力サンプル毎に3つの演算だけであり、これは実際、入力サンプル毎に4回の演算を必要とする式(9)の時間領域対応部分のより少ないことに留意されたい。しかし、時間領域スプライン係数を得るためには、IDFTを $Q(K)$ に適用する必要がある

$$(8)$$

アプローチの1つはフィルタリングを順方向（因果的）演算と逆方向（非因果的）演算とに分割することである。式(8)は、式(8)の補助シーケンス f_n と安定した極すなわち、 $p = 2 - \sqrt{3}$ を使用することによって、容易に次の2つの1次帰納に分割することができる。

$$(9)$$

る。WIデコーダによって処理されるデータはすでにIDFT領域で与えられている。これは信号 $S(t, K)$ である。従って、スプライン変換には $W(K)$ を使用するのが好都合である。また、WI処理のために必要な時間スケールの正規化は、好都合にも単純にゼロを K 番目の軸に沿って $S(t, K)$ に追加することによって行われる。さらに、DFTは有利にも固定した基数2のサイズ N に増加されるので、固定サイズIFFTが有利にも利用できる。このIDFTの結果はサイズ N のスプライン係数シーケンス q_n である。

【0031】本発明の例示としての実施形態の1つによれば、ここで再生音声信号の最終的合成が以下のように行われる。循環関数 $c(t)$ が使用され、正規化サイクル $T = N$ の小数部によってサンプリングの瞬間を探し出す。式(7)に含まれる4つの関連するスプライン係数が識別される。これらの係数は前回のアップデートのスプライン・ベクトルからの対応する係数、すなわち $S(t-1, K)$ から得られたものによって補間される。最後に、式(7)を使用して、数値 $s(t)$ が得られる。この処理は t の十分な数値について繰り返され、出力信号アップデート・バッファを満たすので有利である。 $c(t)$ はアップデートを通じて連続性を維持していることに留意されたい。すなわち、それは前回のアップデートの最後の数値から増加する。しかし、これはモジュロ T の実行であり、基本周期の仮定と一線をなす。

【0032】本発明による複雑さの低いWIコードで使用するための第1の例示としての波形補間処理のブロック図が図3に示される。詳細には、図3に示すWI処理は上記の説明による3次スプラインを使用して波形補間を実行する。すなわち、ブロック31では入力スペクトルにゼロをパディングし、固定した基数2のサイズを確保する。その後、ブロック32では上記で説明したスプライン変換を行い、ブロック33では結果として生じたデータのIFFTを行う。ブロック34は、スプライン係数の補間が現在と過去の波形に基づいて行われる（ブロック38）ように、結果として生じたデータの各集合を保存するために使用される。ブロック36は現在の入力ピッチの数値と（ブロック35で保存された）過去の入力ピッチの数値を演算して動的時間正規化を行い、それに基づいて、ブロック37ではブロック38で補間さ

れるスプライン係数が決定される。最後に、ブロック39で3次スプライン補間が行われ、結果として出力音声波形が(時間領域で)生成される。

【0033】<E. 疑似計数論的スプラインを使用する複雑さの低い波形補間>本発明の他の例示としての実施形態によれば、上記で説明した方法の一変形が、スプライン変換(すなわちスプライン・ウィンドウ)の使用を除去することによって必要な計算をさらに低減する。これは、信号処理技術に熟練した者には周知であり、例えば、上記で示したM. Unser他「B-スプライン信号処理: 第I部-理論」で説明されている計数論的スプラインの概念に基づいている。計数論的スプライン表示は基底関数に1つの追加条件すなわち、ノードで厳密にゼロである($t=n$ かつ $t \neq 0$ の場合 $B(t)=0$)という条件を課すことによって得られる。その結果、それはもはやローカル有限サポートを有さない。しかし、それは、上記で論じた「サイン(x)/ x 」関数と同様、

$$s(t-n) = [(t-n)^3, (t-n)^2, t-n, 1] \begin{bmatrix} -0.75 & 1.25 & -1.25 & 0.75 \\ 1.50 & -2.25 & 1.50 & -0.75 \\ -0.75 & 0.00 & 0.75 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \\ q_{n+1} \\ q_{n+2} \end{bmatrix} \quad (13)$$

ここで $n \leq t \leq n+1$ であるが、これは行列の数値以外は式(7)と同じである。 $t=0$ と設定することによって(行列の一番下の列に留意されたい)入力サンプルとスプライン係数との関係が与えられるが、これは単純に次の式になる。

$$s(n) = q. \quad (14)$$

すなわち、入力サンプルそのものがスプライン係数であるので、それ以上の変換は必要ない。補間器の複雑さは、フィルタリングとウィンドウ処理が有利にも回避される以外は上記で説明した実施形態と同じである。これはサンプル毎に3つの演算を節約し、それによってデコーダの複雑さをさらに低減する。さらに、現在と過去のスプライン係数を保存する追加RAMの必要がなく、スプライン・ウィンドウを保持する追加ROMも必要ないことに留意されたい。

【0034】疑似計数論的スプラインは本来の計数論的スプラインに近似したものにすぎないので、疑似係数論的スプラインに基づくアプローチの性能(すなわち、再生された音声信号の質に関する性能)は、標準3次スプラインに基づくものほどよくないと考えられることに留意されたい。しかし、モデル化と量子化処理の際にデータに追加される重みのレベルは、通常疑似計数論的スプラインによる補間器の使用によって追加される雑音よりはるかに高い。従って、複雑さが低減される利点のほうがこうした近似物を使用する欠点より重要である。

【0035】本発明による複雑さの低いWIコードで使用するための第2の例示としての波形補間処理のプロッ

急速に減衰することに留意されたい。本発明の例示としての実施形態によってここで使用される疑似計数論的スプラインは、他の(すなわち連続性の)条件を緩和することによってこの追加条件を満足させる有限サポート基底関数の使用に基づいている。3次スプラインを使用する上記で説明した場合のように、 $-2 \leq t \leq 2$ のサポートに対する3次対称基底関数を使用される。しかし、次の1つの追加条件が課される。

$$B_1(1) = B_1(-1) = 0 \quad (12)$$

従って、1つの連続性条件だけが放棄される。第2導関数はノード $t=-2$ および $t=2$ で任意の数値を有することが許容される。基底関数とその第1導関数はこれらの点でゼロであることに留意されたい。これらの条件下で基底関数を導き、行列の形で補間演算を表すことによって次の式が与えられる。

【数6】

ク図が図4に示される。詳細には、図4に示すWI処理が、上記の説明による疑似計数論的スプラインを使用して波形補間を実行する。すなわち、図4に示す例示としての波形補間処理の動作は、スプライン変換(ブロック32)が不要になるため除去され、3次スプライン補間(ブロック39)が疑似計数論的スプライン補間(ブロック49)によって代わる点以外は図3に示す例示としての波形補間処理と同様である。

【0036】<F. 複雑さの低い信号分解>上記で記したように、SEW/REW分析は、すべての高調波指標Kについてスペクトル $R(t, K)$ の並列フィルタリングを必要とする。従来のWIコードでは、これは通常20タップ・フィルタを使用して行われる。これは従来技術のWIコードの全体的な複雑さの主要な原因である。詳細には、この処理はコード化して伝送する必要のあるスペクトルの2つのシーケンス、すなわちSEWシーケンスとREWシーケンスを生成する。SEWシーケンスは量子化の前にダウンサンプリングできるが、REWは完全な時間と周波数の解像度で量子化する必要がある。しかし、2.4 kbps以下のコーディング・レートでは、通常のビットの余裕(上記を参照)は小さすぎて、データの有益な表示を生じることができない。この問題の一例として、ピッチ周期80サンプルで、アップデート間隔約12 msecの場合を検討してみよう。通常のフレームサイズを25 msecとすると、各フレーム毎に約2回のアップデートがある。通常、振幅のDFTだけが量子化されるので、1フレーム中量子化される

のは $(80/2) \times 2 = 80$ REW の数値である。しかし、ビットの余裕が許容するのは、REW 量子化器について 6 ビット/フレーム（すなわち、3 ビット/スペクトラム）、すなわち、0.075 ビット/成分だけである。明らかに、この場合伝送できるのは REW の振幅スペクトルの非常に大まかな近似値だけである。実際、上記で示した W. B. Kleijn 他「複雑さの低い補間コード」では、REW 信号は多項式曲線近似技術を使用して徹底的に平滑にされ、わずか 5 パラメータにパラメ

$$SEW(t, K) + REW(t, K) = 1 \quad (15)$$

LP スペクトルの平滑さに関するこの推定は、低レート音声コーディング、詳細には WI によるコードで広範に使用されてきた。これはビットのリソースがない場合に行うには妥当な推定であるが、特に通常の WI コードの場合のように短いフレームについてスペクトルを取る場合には LP スペクトルの著しい表示不足である。そのため SEW 信号と REW 信号は量子化処理で大きく歪められ、コーディングの後原信号の信号特性の多くは残らない。

【0038】 原残差信号の分析（例えば、分解）と WI コーディング環境で実際に行われる量子化映像度との実質上の不整合の存在を認識して、本発明の 1 つの例示としての実施形態は従来技術の WI コードで行われるものより単純な分析を提供する。詳細には、量子化段階で情報の大部分が失われるだけの非常に高解像度の非常に高価な分析を行う必要はないということが認識される。コードの性能は本質的に量子化器によって決定されるので、理論上より単純な分析が使用可能である。従って、本発明の例示としての実施形態によれば、信号分解およびコーディングのタスクに対して新しいアプローチがなされ、SEW と REW が定義され処理される方法を変える。

【0039】 <1. 非構造化成分の複雑さの低い信号分解> 本発明の 1 つの例示としての実施形態によれば、残差信号の非構造化成分は単に適切に整合された正規化電流と以前のスペクトルとの差を得ることによって表される。これは、通常従来 WI コードに見られる 20 次フィルタを 1 次フィルタに代えることによって生じる REW 信号生成の単純化と本質的に同等である。例えば、有声音声では、この差は非構造化ランダム成分を反映する。これをここでは単純にランダム・スペクトル (RS) と呼ぶ。RS は低次 (2 次または 3 次) 直交多項式

$$RS(K) = |S_1(K) - S_2(K)| \quad (16)$$

これは RS の定義と一致している。対応するエネルギーは次の式で表される。

$$E = \sum_K |S_1(K) - S_2(K)|^2 = 2 - 2 \sum_K S_1(K) S_2^*(K) \quad (17)$$

ここで最後の項は対応する時間領域信号間の相互相関の平方として認識される。これらの信号は適切に整列された入力信号（すなわち LP 残差）の 2 つの連続するスナ

ータ化される。

【0037】 同様の状況が SEW 信号についても存在する。通常のビットの余裕（上記を参照）によればフレーム毎に 7 ビットのみが利用可能である。従って、約 800 Hz の SEW ベースバンド・スペクトルのみが通常コーディングされる。それより高い帯域は通常全体に平滑な LP スペクトルを仮定して推定されるので、次の式のようにになる。

展開（例えば、スペクトル毎に 3 または 4 のパラメータを使用する）によって平滑化されるので有利である。これは、どちらのスペクトルもほとんどいつも周波数とともに単調に増大する通常の平滑化された SEW 信号と通常の平滑化された RS を検出することによって見られる。言い換えれば、残差信号は、高い周波数帯域では常に単調に構造化の度合いが少ない。各 RS をコーディングするビットの割当てが 3 ビットだけだと仮定すると（上記の通常のビット割当ての議論を参照されたい）、RS 量子化器によって使用できるのは、こうして平滑化されたスペクトルのうち 8 だけである。

【0040】 平滑化された RS の長いシーケンスに対して従来方法で 3 ビットのベクトル量子化器 (VQ) をトレーニングすることによって、8 つのコードブック・スペクトルの集合が生成される。こうした例示としてのコードブック・スペクトルの集合が図 5 に示される。本発明の例示としての実施形態によれば、平滑化と量子化は、ベクトル毎に 3 つのフルサイズ内積を行うことによって（上記で示した W. B. Kleijn 他「複雑さの低い波形補間コード」で説明されているように）コーディング処理中に結合できる。しかし、例示としてのコードブック・スペクトルの集合の配列が、単純化の度合いを追加していることに留意されたい。詳細には、図 5 に示す曲線が指標とともに単調に増加しているので、それらは、それらのエネルギーと等価であるそれらの下の面積に基づいて、個別に指示できる。発見的に、これは基準化パラメータが、RS コードブックのエントリを指示する入力データから計算できるということを意味している。言い換えれば、コードブックのエントリ（例えば、図 5 の例示としての曲線）は、2 つの整合された正規化スペクトルの振幅の差の平滑化されたバージョン、すなわち次の式を表しているが、

【数 7】

アップショットである。アップデート間隔のサイズがほぼ 1 ピッチ周期である場合、この相互相関は入力のピッチの遅れの相関 $C(P)$ に関連するが、ここ DEP はピッ

チ周期であり、 $C(\cdot)$ は標準相関関数である。従って、(因数 2 を無視すると) パラメータ $u = 1 - (C(P))^2$ が、コードブックの初期の「ソフトな指標」として本質的に使用される。量子化テーブルを使用すると、 u は有利にも、RS 曲線 (すなわちコードブック・エントリ) を指示する範囲 $[0, 7]$ の指標にマップされる。

【0041】上記のアプローチは、エンコードの複雑さの観点から 4 つの主要な利点を有している。第 1 に、明示的な高解像度 RS を生成する必要がない。第 2 に、整合の必要がない。第 3 に、フィルタリングの必要がない。第 4 に、曲線近似の必要がない。しかし、本発明のこの例示としての実施形態によれば、現在のアップデート・レートではピッチの遅れの相関が見られる。

【0042】上記で定義されたパラメータ u は信号中の「無音声」のレベルを反映している。これは常に無声領域では高く、有声領域では低いので、その時間的な強弱はある程度予測可能である。これは VQ をこのパラメータの連続的な数値に適用することによって有効に利用さ

$$I = (\text{seed} = ((++\text{seed}) * 17) \& 4096) >> 5 \quad (18)$$

【0044】 <2 、構造化成分の複雑さの低い信号分解 > 通常の WI コーダでは、SEW 信号は、20 タップ FIR (有限インパルス応答) ローパスフィルタを使用して、時間軸に沿った適切に整合されたピッチサイズ・スペクトルのシーケンスの各高調波成分をフィルタリングすることによって得られる。フィルタリングされたシーケンスはその後フレーム毎に 1 スペクトルに減少される。これはフレーム毎に一度これらのスペクトルの加重平均を取るのと同様である。前に示したように、本発明のある例示としての実施形態によれば、フィルタリングと整合の両方が回避されるので有利である。

【0045】本発明のある例示としての実施形態では、構造化信号が以下のように処理されるので有利である。現在のフレームについてピッチ周期 P を仮定すると、整数 M を含むピッチ周期の新しいフレームが決定される。通常、新しいフレームは名目のフレームに重なる。その後、ここで AS と呼ぶピッチサイズ平均スペクトルが、DFT をこのフレームに適用し、MP サイズ・スペクトルを因数 M によってデシメイトし、その結果を正規化す

$$S(K) = A(K) - R(K) \quad (19)$$

【0047】WI コーダのビットの余裕は、上記で説明したように AS のコーディングのために 7 ビットしか提供しない。LP 残差のもっと低い周波数が知覚的にはより重要なので、本発明の例示としての実施形態によれば、SEW スペクトルの下の 20 % を含むベースバンドだけがコード化されるので有利である。AS の振幅スペクトルの残りの部分は、例えば、平坦であり、 $A(K) = 1$ であると推定される。

【0048】従って、例示としての複雑さの低いコーダは AS ベースバンドをコーディングし、コーディングさ

れる。従って、本発明の他の例示としての実施形態では、ベクトル毎に 3 ビットを使用して RS を直接量子化する代わりに、6 ビット VQ が有利にも使用され、フレーム内の u ベクトルを量子化し伝送する。受信機では、デコーディングされた u の数値が直交多項式パラメータの集合にマップされ、そこから平滑化された RS スペクトルが生成される。

【0043】デコーディングされた RS が振幅スペクトルを表すことに留意されたい。本発明の例示としての実施形態によれば、完全な複合 RS は、非構造化信号の仮定と一致するランダム位相スペクトルを追加することによって得られる。ランダム位相は、例えば、位相テーブルをランダム・サンプリングすることによって安価に得られる。こうした例示としてのテーブルは 128 の半径 1 の 2 次元ベクトルを保持する。 $0 < I < 128$ であるとき、この表の指標 I は、例えば、有利にも高速ビット演算によって実現される次の式の C 言語指標再帰によって疑似ランダムに生成される。

ることによって得られる。このアプローチはスペクトル整合の必要を除去するので有利である。DFT の複雑さを低減するために、SEW フレームはまず基数 2 のサイズ $N > MP$ にアップサンプリングされ、その後高速フーリエ変換 (FFT) が使用される。この時間正規化は依然として MP に等しいスペクトルのサイズに影響しないことに留意されたい。アップサンプリングは、例えば、上記で説明した 3 次スプライン補間を使用して行われる。

【0046】平均スペクトル AS は、簡単なフィルタを使用して、SEW の単純化されたバージョンとして見ることができる。従来の WI コーダによって生成した REW および SEW 信号と異なって、 $A(K)$ と (非平滑化) $R(K)$ は、2 つの相補的フィルタによって生成されたものではないので、相補的ではない。実際、 $A(K)$ 自体は、LP の振幅スペクトルの現在の推定値として見られる。従って、構造化スペクトル (SS) と考えられるスペクトルの一部は、次の式のようなになる。

れた結果を 1 フレーム毎に 1 回伝送する。コーディングは例示としては、 D が 0、 $2 * P / 2$ または 10 の低い方である変数次元 D の 10 次元 7 ビット VQ を使用して行われる。 $D < 10$ の場合、コードベクトルの最初の D 項だけが使用される。受信機では、AS ベースバンドが合成アップデート・レートで補間され、そこから $S(K)$ スペクトルが計算される。

【0049】振幅スペクトル $SS(K)$ は周期的信号を示す。従って、固定位相スペクトルが有利にもそこに加えられ、自然音声で観察されるようなあるレベルの位相

分散を提供する。これによって周期性が維持されるとともにブザー音が避けられる。現実の話者から得られる位相スペクトルは、例示としては半径 1 の 64 の複素数値を有する。それは RS (最初の 64 エントリ) によって使用される同じ位相テーブルに保持されるので、追加 ROM を必要としない。結果として生じる複合 SS は例示として複合 RS と結合され、現在のアップデートに関する最終量子化 LP スペクトルを形成する。

【0050】<G. アップデート・レートの検討>従来の WI コーディングでは、SEW と REW は、現在のピッチとは無関係に、任意の望ましいアップデート・レートで生成・処理される。さらに、レートはエンコーダとデコーダで異なることもある。固定レート (例えば、2.5 msec のアップデート間隔) が使用される場合、データ流れ制御は直接的である。しかし、スペクトルのサイズは実際にはピッチに依存し、結果として生じる計算負荷も同様である。従って、固定アップデート・レートでは、複雑さはピッチ周期の数値とともに増大する。最大計算負荷が問題になることが多いので、複雑さを「均等化」することが有利である。従って、本発明の例示としての実施形態によれば、ピーク負荷を低減するために、アップデート・レートが有利にもピッチ周波数と比例して変更される。

【0051】通常の従来の WI コーダの場合、短時間スペクトル・スナップショットがピッチサイクル間隔で処理されることに留意されたい。これはほぼ周期的な音声の場合、ピッチレートで信号の強弱を監視すれば十分であるという仮定に基づいている。こうした可変サンプリングは SEW/REW 信号フィルタリング段階である種の困難を提起するので、ある特殊なフィルタリング処理が必要となる。

【0052】しかし、本発明による例示としての複雑さの低い WI (LCWI) エンコーダでは、固定サイズ FFT を使用してフレーム毎に一度 AS が処理されるので、こうした困難は存在しない。RS は固定レートでアップデートされる一方でピッチ間隔 (すなわちピッチの遅れの相関) の変化を測定する u パラメータによって表される。

【0053】従来の WI デコーダと例示としての LCWI デコーダの両方で、アップデート・レートはピッチ依存なので、負荷を均等化し、結果が過度に周期的 (すなわち、レートが遅すぎる) にならないようにする。さらに、例示としての LCWI コーダのスプライン変換と IFFT は、ピッチの数値をもっとも近い基数 2 の数に切り上げることによってピッチ依存になる。これはピッチ範囲による計算負荷の変化を低減するので有利である。従って、現在のピッチを仮定すると、アップデート・レート制御 (URC) 処理が有利にも利用され、スペクトルが再生され、出力信号が補間される合成サブフレーム・サイズを決定する。u パラメータは例示としては固定

レートで (例えば、フレーム毎に 2 回) 送信されるので、より高いアップデート・レートが必要な場合デコーダで補間される。

【0054】<H. LP パラメータの複雑さの低い量子化>例示としての LCWI コーダでは、複雑さの低いベクトル量子化器 (LCVQ) が LP パラメータをコーディングする際使用され、さらに計算負荷を低減する。例示としての LCVQ は、ここに記載されるのと同様に、引用によって本明細書の記載に援用する、J. Zhou 他「線形スペクトル周波数の単純な高速ベクトル量子化」、ICSLP'96 会報、第 2 巻、945~948 ページ、1996 年 10 月で詳細に説明されているものに基づいている (ここで説明されている例示としての LCVQ は必ずしも WI コーダに特定のものではなく、有利にも他の LP による音声コードでも使用できることに留意されたい)。

【0055】例示としての LCVQ では、LP パラメータは 10 の線形スペクトル周波数 (LSF) の形態で与えられる。10 次元 LSF ベクトルは、1.2 kbps コーダでは 30 ビット、また 2.4 kbps コーダでは 25 ビットを使用してコーディングされる。フルサイズの 25 または 30 ビットの VQ は実際には実現可能でないもので、LSF ベクトルは普通 3 つの下位ベクトルに分割される。詳細には、3 つの LSF 下位ベクトルのサイズは、1.2 kbps コーダの場合 (3, 3, 4)、また 2.4 kbps コーダの場合 (3, 4, 3) である。3 つの下位 VQ に割り当てられたビット数はそれぞれ (10, 10, 10) および (10, 10, 5) である。各下位 VQ はフル探索 VQ を含むが、これは包括的な探索が 1024 (または 32) のコードベクトルの候補にわたって行われることを意味する。しかし、本発明による例示としての LCWI コーダでは、フル探索 VQ は以下で説明されるより高速な VQ によって置き換えられる。

【0056】すなわち、ここで使用される例示としての高速 VQ は、フル探索 VQ より約 4 倍高速である。これは同じ最適にトレーニングされたコードブックを使用し、同じ性能レベルを達成する。詳細には、これは当業技術分野に熟練した者には周知の分類 VQ の概念に基づいている。主コードブックは多数の下位コードブック (クラス) に分割される。入力ベクトルはまずあるクラスに属するものとして分類される。その後そのクラスとその隣にある少数のクラスだけが探索される。分類段階はまた別の小さいサイズの VQ によって実行されるが、その VQ のエントリはそれら自身のクラスを示す。このコードブックは有利にも主コードブックの中で実現されるので、コードベクトルのための追加メモリ・ロケーションは必要ない。しかし、クラスのポイントを保持するために合計メモリのわずかな増加 (約 2%) が必要である。

【0057】<1. 例示としての複雑さの低いW I コーダ>図6は、本発明の1つの例示としての実施形態によるL C W I コーダのブロック図を示す。すなわち、図6は、その例示としてのブロック図とともにエンコーダ61を示し、その例示としてのブロック図とともにデコーダ62を示し、エンコーダとデコーダの間の例示としてのデータの流れを示す。詳細には、伝送されるビット・ストリームには例示としてはそれぞれG、L、R、AおよびPで示される量子化利得、L S F、R S、A S、およびピッチの指標が含まれる。

【0058】<1. 例示としてのL C W I エンコーダ>図6に示される例示としてのエンコーダでは、L P 分析が入力音声に適用され(ブロック6104)、上記で説明したL C V Q が使用されてL S F をコーディングする(ブロック6109)。入力音声利得がブロック6103でフレーム毎に4回の固定レートで計算される。利得は、主フレーム内で均一に間隔の開いた重なり合うピッチサイズの下位フレームのR M S として定義される。これによって変化のない有声音声では利得の輪郭が非常に平滑になる。ピッチサイクルが短すぎる場合、2つかそれ以上のサイクルが使用される。これによって重要かもしれない利得のキューのセグメントを飛ばすことが防止される。4つの利得はフレーム毎に1つの利得ベクトルとしてコーディングされる。例示としての2. 4 k b p s バージョンのエンコーダの場合、10ビットが利得に割り当てられる。利得ベクトルは「超利得」と呼ばれるそのR M S の数値によって正規化される。2段階L C V Q が使用される(ブロック6109)。まず正規化ベクトルが6ビットV Q を使用してコーディングされる。その後、超利得の対数(log)が4ビット量子化器を使用して差動的にコーディングされる。このコーディング技術によって量子化器のダイナミックレンジが増大し、同時に、例えば、有声音声の開始を表す利得の短期間の(すなわち1つのベクトル以内の)変化が表せるようになる。例示としての1. 2 k b p s バージョンのエンコーダでは、超利得は使用されず、単一の8ビット4次元V Q が対数利得に適用される。

【0059】入力がL P 係数を使用して逆フィルタリングされ、L P 残差が得られる(ブロック6101)。残差に対してピッチの検出がなされ、現在のピッチ周期が得られる(ブロック6102)。R S およびA S 信号が上記で説明したように処理される。ブロック6105では、u 係数が生成され、ブロック6110では、u 係数は例示としての1. 2 k b p s コーダでは5ビット、また2. 4 k b p s コーダでは6ビットをそれぞれ使用して2次元V Q によりコーディングされる。例示としての2. 4 k b p s コーダでは、A S ベースバンドが7ビットを使用して10次元V Q によりコーディングされる(ブロック6106、6107、6111および6112)。1. 2 k b p s コーダでは、A S は処理・コーデ

ィングされず、むしろ定数、すなわちすべてのKについて $AS(K) = 1$ と考えられる。従って、図6のブロック6106、6107、6111および6112は例示としての1. 2 k b p s コーダでは存在しない。

【0060】<2. 例示としてのL C W I デコーダ>図6に示される例示としてのデコーダでは、受信されるピッチの数値がブロック6209のアップデート・レート制御(U R C)で使用され、現在のアップデート・レート、すなわち、補間および合成処理全体が行われる下位フレームの数を設定する。ピッチはブロック6205で前の数値を使用して補間され、数値が各下位フレームに割り当てられる。

【0061】ブロック6201では、超利得が差動的にデコーディングされ、べき乗される。正規化された利得ベクトルがデコーディングされ、超利得と結合される。またU R C によって要求される場合、4つの利得の数値がより長いベクトルに補間される。L P 係数がフレーム毎に一度デコーディングされ、前のものによって補間されてU R C が必要とする数のL P ベクトルが得られる(ブロック6202)。L P スペクトルが、D F T 6206をL P ベクトルに適用することによって得られる。入力が10サンプルだけであるので、これは有利にも複雑さの低いD F T であることに留意されたい。D F T は再帰的に行われ、高価な三角関数を避ける。また、F F T が3次スプラインによる再サンプリングと共に使用されることがある。

【0062】ブロック6203では、U R C が必要とする場合、R S ベクトルがデコーディングされ補間される。各uの数値が拡張パラメータの集合にマップされ、平滑化された振幅のR S が生成される(ブロック6207)。ランダム位相がブロック6210で付加され、複合R S を生成する。

【0063】例示としての2. 4 k b p s コーダでは、A S がデコーディングされ、前のベクトルによって補間される(ブロック6204)。S S 振幅スペクトルが、ブロック6208でR S を引き算することによって得られ、その後S S の位相がブロック6211で加算される。複合R S およびS S データが結合され(ブロック6213)、その結果がL P スペクトルによって形成され、利得によって基準化される(ブロック6212)。その結果が、波形補間モジュールに適用され(ブロック6214)、コーディングされた信号が出力される。波形補間モジュールは、図3の例示としての波形補間処理か、図4の例示としての波形補間処理か、または本発明の原理による他の波形補間処理を含む。

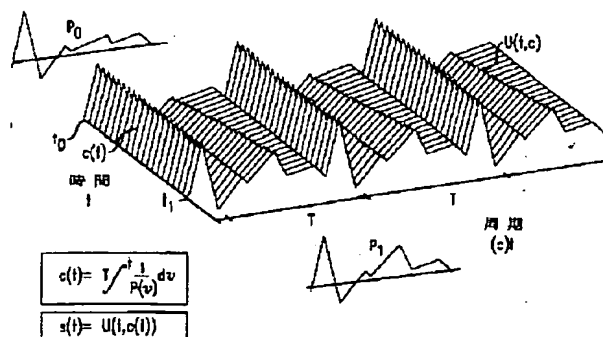
【0064】最後に、(好適には軽度の)事後フィルタリングがブロック6215で適用され、出力コーディング雑音を再形成する。例えば、J. H. Chen 他「コーディング音声の質を向上するための適応事後フィルタリング」、米国電気電子学会音声・音響処理会報、第3

巻、1995年59～71ページで説明されているものと同様のLPによる事後フィルタが使用される。こうした事後フィルタはLPフォーマットのパターンを向上させ、それによってホルマント間の雑音を低減する。また、事後フィルタリング操作は、上記で示したW. B. Kleijn他「複雑さの低い波形補間コード」で説明されているWIコードでなされるように、LP形成段階（すなわち、ブロック6212）に含まれることもある。しかし、3次スプライン補間器の雑音を含む全体的な雑音を低減するために、事後フィルタは好適には、図6の例示としての実施形態に示すように、合成処理の最後に配置される。

【0065】＜J＞追補＞説明をわかりやすくするために、本発明の例示としての実施形態は独立した機能ブロック（「プロセッサ」と名付けられた機能ブロックを含む）を含むものとして示される。これらのブロックが表す機能は、ソフトウェアを実行できるハードウェアを含むがそれに制限されない共用または専用のハードウェアの使用を通じて提供される。例えば、ここに示されるプロセッサの機能は、単一の共用プロセッサまたは複数の独立したプロセッサによって提供される。さらに、ここでの「プロセッサ」という術語の使用は、ソフトウェアを実行できるハードウェアにのみ関するものと解釈されるべきではない。例示としての実施形態は、Lucent Technologies社のDSP16またはDSP32Cといったデジタル信号プロセッサ（DSP）ハードウェア、以下論じる演算を行うソフトウェアを保存するための読み出し専用メモリ（ROM）およびDSPの結果を保存するためのランダムアクセス・メモリ（RAM）を含む。超大規模集積（VLSI）ハードウェア実施形態が、汎用DSP回路と協力するカスタムVLSI回路とともに提供される。任意の、およびすべてのこれらの実施形態はここで使用される「プロセッサ」という語の意味に含まれるものと考えられる。

【0066】本発明の多数の特定の实施形態がここで示

【図1】



され説明されたが、これらの実施形態は単に、本発明の原理を適用する際に考案される多くの可能な特定の装置の例にすぎないことが理解されるべきである。非常に多くの多様な装置が、当業技術分野に普通に熟練した者によって本発明の精神と範囲から離れることなく本発明の原理に従って考案できる。例えば、添付の請求項で使用される「信号受信機」、「スプライン係数発生器」、「信号シンセサイザ」といった術語の使用は、相応じて識別される機能を行う任意の装置を対象にすることを目的とするのであって、3.5 U. S. C.、112節、6項に規定されているような「手段プラス機能」の形態にあるものと考えられるべきではない。同様に、これらの請求項の要素は、それらの範囲内で、「明細書とそれに相当するものの中で説明された... 対応する構造」に制限されることを目的とするものではない。しかし、そのように制限されるものではないが、本発明の実施形態の1つの組み合わせは、実際は本明細書とそれに相当するものの中で説明された構造を含むことに留意すべきである。

【図面の簡単な説明】

【図1】波形補間コードによって有利にも発生する一連の平滑に変化する波形を含む曲面を示す図である。

【図2】従来の波形補間コードのブロック図を示す図である。

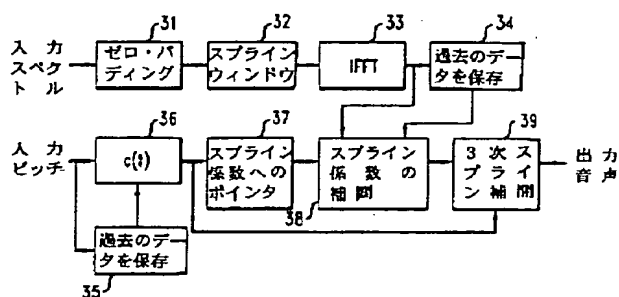
【図3】本発明の第1の例示としての実施形態による3次スプライン表示に基づいた波形補間のブロック図を示す図である。

【図4】本発明の第2の例示としての実施形態による疑似係数論的スプライン表示に基づいた波形補間のブロック図を示す図である。

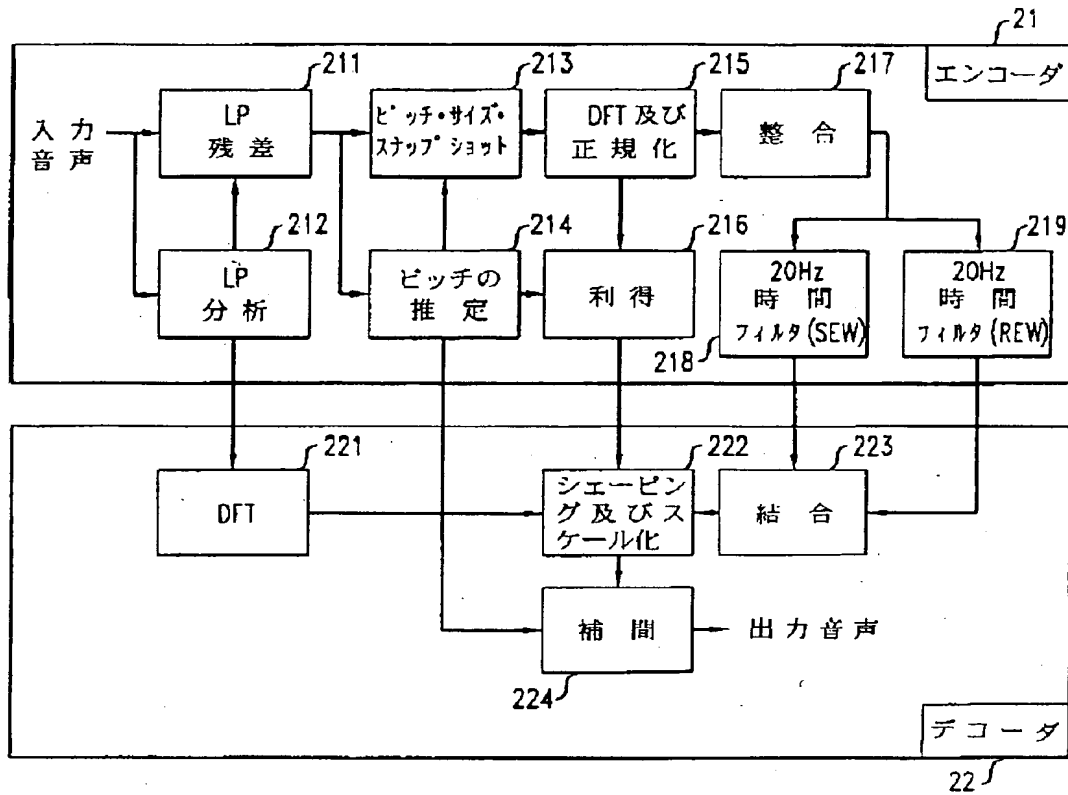
【図5】波形補間コードのランダム・スペクトル・コードブックのための平滑なスペクトルの例示としての集合を示す図である。

【図6】本発明の例示としての実施形態による複雑さの低い波形補間コードのブロック図である。

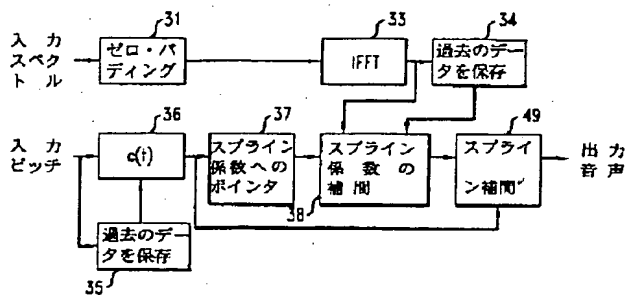
【図3】



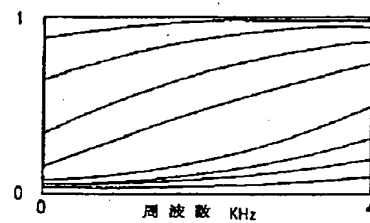
【図 2】



【図 4】



【図 5】



【 図 6 】

